Amendments to the Specification

The paragraph beginning on page 5, line 27, is amended as follows:

FIG. 9A[a] illustrates a synthetic protocol for a diamine virtual combinatorial library that was used to demonstrate the effectiveness of the present invention.

[The entire text on page 6 is amended as follows:]

FIG. 9B [b] illustrates a synthetic protocol for a Ugi reaction that was also used to demonstrate the effectiveness of the present invention;

FIG. 10A[a] shows a query structure that was used to demonstrate the effectiveness of the present invention;

FIG. 10B[b] shows another query structure that was used to demonstrate the effectiveness of the present invention;

FIG. 11A[a] is a graph that illustrates a similarity profile of the diamine virtual combinatorial library of FIG. 9A[a];

FIG. 11B[b] is a graph that illustrates a similarity profile of the Ugi virtual combinatorial library of FIG. 9B[b];

FIGS. 12A [a] and 12B[b] are graphs that illustrate the overlap between stochastic selections, generated using the present invention, and reference selections;

FIG. 13 is a table that summarizes the experimental results obtained when using an embodiment of the present invention;

FIGS. 14A[a], 14B[b], 15, 16A[a], 16B[b] and 17 are graphs that show how the selection of particular variables affects experimental results of the present invention;

FIG. 18 shows the structures of some of the most similar compounds found during experimental testing of the present invention;

FIG. 19 shows an exemplary environment in which the present invention can be used;

- 4 -

Lobanov *et al.*
Appl. No. 09/506,741

FIG. 20 shows an example of a computer system that can be used to implement the present invention;

FIG. 21 shows examples of diamines, and acidchlorides and halocarbons (i.e., alkylating/acylating agents), associated with the diamine virtual combinatorial library that was used to demonstrate the effectiveness of the present invention; and

The paragraph beginning on page 9, line 17 is amended as follows:

In contrast, using the present invention, a similarity evaluation of the same 6.75 million possible compounds, using the same dual processor 400 MHZ Intel Pentium II machine, required only 30 minutes. This large reduction in time is due to the fact that the present invention does not perform enumeration, characterization, and similarity evaluation for all of the 6.75 million possible compounds.

The paragraph beginning on page 36, line 24 is amended as follows:

The effectiveness of the present invention has been demonstrated by the inventors through experiments using two different virtual combinatorial libraries. The first was a diamine virtual combinatorial library, generated by combining a diamine core with a set of alkyl halides or acid chlorides, as shown in FIG. 9A[a]. The structure 902a represents a diamine; R1-X and R2-X each independently represent an alkyl halide or acid chloride. Although physical synthesis of this library could prove problematic (the synthetic sequence involves selective protection of one of the amines and introduction of the first side chain, followed by deprotection and introduction of the second side chain), for the purpose of a study it can be assumed that one of the amino groups on the diamine core reacts with the first reagent, while the other reacts with the second reagent. A substructure search in the Available Chemicals Directory (ACD) yielded 1,036 suitable diamines and 826 alkylating/acylating agents. These reagents were used to generate a virtual combinatorial library associated with over 706 million (1036 ′ 826 ′ 826) possible products (i.e.,reagent combinations). Since descriptors for the fully enumerated virtual combinatorial library could not be computed in a timely fashion, and since for validation purposes the inventors needed

- 5 -

Lobanov *et al.*
Appl. No. 09/506,741

to compare their results with conventional selections from a fully characterized library, a smaller 6.75 million-membered library (i.e., a virtual combinatorial library associated with 6.75 million reagent combinations) was produced by choosing 300 diamines and 150 alkylating/acylating agents at random. Hereafter, the term "diamine virtual combinatorial library" will refer to this smaller library, unless noted otherwise. Examples of diamines, and acidchlorides and halocarbons (i.e., alkylating/acylating agents), associated with this smaller library are shown in FIG. 21.

[The paragraph beginning on page 37, line 20, is amended as follows:]

The second virtual combinatorial library was based on the Ugi reaction, and involves an organic acid (R1-COOH), an amine (R2-NH$_2$), an aldehyde (R3-CHO) and an isonitrile (R4-CN), as shown in FIG. 9B[b]. A substructure search in the ACD yielded 1,681 suitable acids, 594 suitable amines, 37 suitable aldehydes, and 17 suitable isonitriles. These reagents were used to build a virtual combinatorial library associated with over 628 million possible compounds (1681 ′ 594 ′ 37 ′ 17). Again, for validation purposes a smaller 6.29 million-membered library (i.e., a virtual combinatorial library associated with 6.29 million reagent combinations) was produced by choosing a random set of 100 acids and 100 amines. Hereafter, the term "Ugi virtual combinatorial library" will refer to this smaller library, unless noted otherwise. Examples of acids, amines, aldehydes, and isonitriles associated with this smaller library are shown in FIG. 22.

[The paragraph beginning on page 38, line 3 is amended as follows:]

First, every reagent combination (approximately 6.75 million) associated with the diamine virtual combinatorial library was enumerated to produce a full set of enumerated compounds. The similarities of the enumerated compounds to a query structure (an antiarrhythmic agent), shown in FIG. 10A[a], were then evaluated. The evaluation of molecular similarity was based on a standard set of 117 topological descriptors computed using a C++ descriptor generation class from the DirectedDiversity® API toolkit, available from 3-Dimensional Pharmaceuticals, Inc.,

Exton, Pennsylvania. The descriptors included a well-established set of topological

indices with a long, successful history in structure-activity correlation such as

molecular connectivity indices, kappa shape indices, subgraph counts, information-

theoretic indices, Bonchev-Trinajstis indices, and topological state indices. These

indices are discussed in detail in Hall et al. "The Molecular Connectivity Chi Indexes

and Kappa Shape Indexes in Structure-Property Relations," Reviews of

Computational Chemistry, Chap. 9, pp 367-422, eds. Donald Boyd and Ken

Lipkowitz, VCH Publishers, Inc. (1991), and Bonchev et al. "Information Theory,

Distance Matrix, and Molecular Branching," J. Chem. Phys. 1977, 67, pp 4517-4533,

both of which are incorporated herein by reference in their entirety. The calculated

descriptors were normalized, and decorrelated using principal component analysis.

The principal components which accounted for 99% of the total variance in the data

(typically 25-30 principal components) were used to define the similarity space.

Pairwise dissimilarity scores were calculated as Euclidean distances between the

vectors associated with the respective compounds in the space defined by the selected

principal components. A higher dissimilarity score indicates compounds that are less

similar to each other, that is, more distant from each other in the principal component

space.

        The paragraph beginning on page 38, line 29 is amended as follows:

        Based on calculated dissimilarity scores, a similarity profile 1102a, shown in

FIG. 11A[a], of the diamine virtual combinatorial library was obtained by counting

the number of compounds falling in each similarity bin 1104a. According to the

distribution, the majority of compounds had dissimilarity scores higher than 4.0.

Next, the 100 most similar compounds with the lowest dissimilarity score were

selected, and were used as a reference to compare all subsequent similarity selections

drawn from the diamine virtual combinatorial library. These reference compounds

represent the absolute best similarity selection of that size which can be obtained from

the diamine virtual combinatorial library using the prescribed descriptors, similarity

measure, and query structure.

The paragraph beginning on page 39, line 10 is amended as follows:

Likewise, every reagent combination (approximately 6.29 million) associated with the Ugi virtual combinatorial library was also enumerated and the similarities of the enumerated compounds to the query structure (a 1.4 µM thrombin inhibitor), shown in FIG. 10B[b], were evaluated. Although the enumerated compounds associated with the Ugi virtual combinatorial library exhibited a more sharp similarity distribution, as shown by the similarity profile 1102b in FIG. 11B[b], the vast majority of the compounds in that library also had dissimilarity scores higher than 4.0. Again, the 100 most similar compounds were identified and were used as a reference to compare subsequent similarity selections from that library.

The paragraph beginning on page 39, line 20 is amended as follows:

Next, the selection method of the present invention, outlined in the discussion of FIG. 5 above, was employed to select the 100 highest-scoring compounds from the diamine combinatorial virtual library based on their similarity to the same query structure of FIG. 10A[a]. Initially 100,000 reagent combinations were selected at random from the virtual combinatorial library (Step 104, Fig. 5) and enumerated (Step 106). Descriptors were calculated for the enumerated compounds (Step 502), pairwise similarity to the query structure was evaluated (Step 504), and the enumerated compounds were ranked based on their similarity (Step 506). The 100 highest-ranking compounds were selected (508) and deconvoluted to produce lists of "preferred" reagents (Step 110). The list of "preferred" reagents was then used to produce the "focused" library (Step 112) and all reagent combinations associated with that "focused" library were enumerated (Step 114). Descriptors were calculated for the enumerated compounds (Step 510), pairwise similarity to the query structure was evaluated (Step 512), and the enumerated compounds were ranked based on their similarity (Step 514). The 100 highest-ranking (most similar) compounds were then selected (Step 516) from the fully enumerated "focused" library based on their similarity to query structure of FIG. 10A[a].

The paragraph beginning on page 40, line 12 is amended as follows:

The dissimilarity scores and identities of the selected 100 compounds derived using the present invention were compared with the dissimilarity scores and identities of the reference selection derived from the fully enumerated library. Based on the average dissimilarity scores of the selected compounds (1.37 vs 1.30), the two selections (i.e., the selection using the present invention and the reference selection) were quite comparable. In fact, as shown in FIG. 12A[a], which shows the overlap between stochastic selections 1202a, 1204a, 1206a, and reference selection 1218a, most of the compounds in the reference set were also found in the stochastic selection. After repeating the stochastic procedure two more times with different random seeds and combining the results, the overlap with the reference selection rose to 96 out of 100 compounds, as shown at 1208a of FIG. 12A[a].

The paragraph beginning on page 40, line 24 is amended as follows:

The same procedure was also applied to the Ugi library and an even better overlap with the reference selection was achieved, as shown in FIG. 12B[b].

The paragraph beginning on page 41, line 6 is amended as follows:

For comparison, as shown in row 1306 of Table 1302, when three independent screens of 200,000 random reagent combinations relating to the diamine virtual combinatorial library were selected and enumerated, a total of only 9 (or 3 on average) out of the 100 most similar structures were retrieved. (This is also shown in FIG. 12A[a] at 1210a, 1212a, 1214a and 1216a.) This result is not surprising since 600,000 compounds constitute approximately 10% of the entire 6.75 million-member diamine library.

The paragraph beginning on page 43, line 29 is amended as follows:

Since the present invention is most useful when applied to massive virtual combinatorial libraries that are intractable by other means, the inventors derived a series of selections from the full diamine virtual combinatorial libraries and Ugi

virtual combinatorial libraries containing 706 and 628 million possible enumerated compounds, respectively, using the same query structures of FIG. 10A[a] and FIG. 10B[b], and varying the same selection parameters (i.e. the size N of the initial pool and the number M of highest-ranking compounds used to derive the focused library). As before, each combination of parameters was tested three times starting from a different random seed, and the results were averaged. The results are summarized in FIG. 17.